

DATA ANALYSIS: GENERALISATIONS OF THE LOCAL APPROXIMATION METHOD BY SINGULAR SPECTRUM ANALYSIS

A.Loskutov, I.Istomin and O.Kotlyarov

Physics Faculty, Moscow State University, Moscow 119899 Russia

Abstract

We study forecasting capabilities of the methods of Singular Spectrum Analysis (SSA) and Local Approximation (LA). A practical implementation of these methods to several time series is described. Details of the algorithms of these methods are discussed. Advantages and disadvantages of SSA and LA are given. On the basis of our results we generalize LA and SSA and propose a new way for time series forecasting including strongly noisy signals. This allowed us to extend the range of application of LA in comparison with the known standard scheme. For the problem of forecasting, the accuracy enhancement of the numerical computations is discussed.

1 Introduction

Analysis of results of any observable is often based on the processing of initial experimental data. In many cases these data represent a time series, i.e. a certain sequence with elements in a chronological order. The processing of the time series with the purpose of extraction of the useful information about properties of the corresponding system is a very important and interesting task. However, in many cases attention is paid not to research of properties of the system, but to the forecast its further dynamics. For example, in meteorology the practical interest refers first of all to the weather forecast in a near future. Thus, besides the study of the system properties, the task of the forecasting of the further trajectory is also of practical significance. This problem does not exclusively belong to meteorology, it also is very important both in geophysics and astrophysics. It concerns the prediction of earthquakes, forecast of solar activity. In financial analysis it concerns the forecast of share prices, exchange rates etc.

The most widespread methods of the forecast of irregular (quasi periodic and stochastic) time series are currently the methods of ARMA type (Auto Regressive Moving Average) [1],

which have already for a long time been applied in statistics, economy and meteorology [2]. The basic idea is expression of the following elements in the time series by the previous ones. This is probably a unique way, which can be used in the situation when any other modeling of the system is not available, except for investigations the observation data. The validity of such an approach was proven only after the Takens paper at the beginning of the 80s [3].

Besides the validation of ARMA, the Takens ideas initiated the development of new methods of analysis and forecasting the time series by the theory of dynamical systems. One of such methods is a method of local approximation (LA), offered for the first time in the work [4] for the forecast of the chaotic time series. This method, which algorithm is briefly described in the present work, has a number of advantages against a traditional method of autoregression. However it is not yet widely spread basically because of difficulties in application to short and noisy time series.

In the present paper we generalize the efficiency of LA for forecasting noisy time series by means of a preliminary filtration of the data with the help of Singular Spectrum Analysis (SSA) [5]. The SSA method is also developed within the framework of nonlinear dynamics, however it is used mainly for definition of the basic components and suppression of noise [6], though there exist algorithms, based on it, which allow the forecast [7]. We hope that the offered generalization of the LA method should admit considerably to expand the area of its application and to the forecast of the chaotic and quasi-periodic behavior of nonlinear dynamical systems.

2 Methods of processing of time series

In this section we give some basis of the time-series analysis. This allows us to generalize the LA method and propose a new way for forecasting chaotic data.

2.1 Method of delays

A foundation of the majority of approaches related to the processing of time series $\{x_1, \dots, x_N\}$ is the construction of a set of delayed vectors (or vectors in the state space) $z_i = (x_i, x_{i+1}, \dots, x_{i+m})^T$, $i = 1, 2, \dots, N - m + 1$ [8]. This is the first and necessary step in the methods of the analysis of time series developed in the framework of nonlinear dynamics. In a certain sense, the state space is equivalent to the phase space of the corresponding nonlinear dynamical system, which generated the time series [3, 9]. Thus, it has been proven that many-dimensional systems can be described by time series, i.e. a scalar function of the system state obtained in an experiment. In its turn, the possibility of the description and

reconstruction of the system dynamics allows, under the certain conditions, to predict its further behavior [9].

As a rule, for the construction of vectors z_i the method of delays is used. This method similarly autoregression, establishes the transformation from an initial one-dimensional (scalar) time series to a many-dimensional (vector) representation. For this transformation, each many-dimensional vector is formed from some number (say m) of successive values of the initial time series. The result can be presented as a set of "photos" of a series made through a window sliding along the series, in which only m consecutive elements may observe simultaneously:

$$X = \left(\begin{array}{ccc} \vdots & \vdots & \\ f_{m+1} & f_{m+2} & \\ \left[\begin{array}{c} f_m \\ \vdots \\ f_2 \\ f_1 \end{array} \right] & \left[\begin{array}{c} f_{m+1} \\ \vdots \\ f_3 \\ f_2 \\ f_1 \end{array} \right] & \dots \left[\begin{array}{c} f_N \\ \vdots \\ f_{N-m+2} \\ f_{N-m+1} \\ f_{N-m} \\ \vdots \end{array} \right] \end{array} \right). \quad (1)$$

\downarrow

\downarrow

Here f_1, f_2, \dots, f_N are values of elements in the series at time $t = 1, 2, \dots, N$. Each square bracket is a vector in m -dimensional space of delays; the sequence of such vectors gives an observation matrix $X_{m \times (N-m+1)}$, where N is the number of elements of an initial series. This matrix, in every column of which there are parts of the same series moved relative to each other, is many-dimensional form of an initial scalar series in space of delays.

In a discrete case the described many-dimensional representation is given by one parameter. This parameter is the dimension of the space of delays, or the embedding dimension m . As shown, in many cases the opportunity of the exact and reliable forecast depends on its correct choice. From the theory of dynamical system it follows that $m \geq 2d + 1$, where d is the attractor dimension of the system, which generated the series. However this condition is not constructive in a choice of the embedding. The most popular algorithm for an estimation of the embedding dimension (and the system dimension) is a Grassberger-Procaccia algorithm [10], but it is also not so efficient for short (up to 10^4 elements) time series.

For periodic and quasi-periodic time series there is an empirical rule according to which it is necessary to choose $m \in (T/5, T)$, where T is an average period [6].

The specified complexities in the determination of the embedding dimension can be overcome by means of its definition within the framework of a used forecast method. In this case an available series is divided into two unequal parts. One of them (the smaller) is used for the quality check of the forecast made on the basis of the other. The dimension, where

the forecast turns out to be the best, is considered optimal for the given series.

2.2 Method of local approximation (LA)

Today the methods of autoregression (in the class of ARMA methods [1]) are most frequently used for forecasting the time series. The autoregression model of the order p -AR (p) has the following form:

$$f_t = a_0 + a_1 f_{t-1} + a_2 f_{t-2} + \dots + a_p f_{t-p}. \quad (2)$$

In this case, to predict the further trajectory of a series, it is necessary first to determine the order of autoregression, and then on the basis of the available data to obtain estimations of autoregression coefficients $\{a_0, \dots, a_p\}$. Here, however there is no unique algorithm of choice of the order, and as a rule, it is chosen due to the type of autocorrelation function. The corresponding coefficients are estimated for all available data and assumed to be constant for any t . Thus, the given approximation is a global one.

In general, the methods of global approximation give a quite good description of the function when the number of free parameters are enough. However, if the function is rather complex, there is no guarantee that we can find such representation which allows to approximate efficiently the analyzed function. In this way we can obtain an exclusive circle: The more complex function, the more parameters are necessary, the more parameters should be estimated, the more data are required. Increase of the number of the used data (which is not always accessible) for a quite complex function requires introduction of additional parameters to the model.

To avoid this cycling one can use LA. Its basic idea is to divide the domain of the function into some local subdomains, construct approximate models and estimate parameters separately in each area. If the function is smooth then the subdomains can be small enough such that the function in each of them does not change too sharply. It allows to apply in each domain simple models (say, a linear one). The main condition of the efficient application of LA is the correct choice of the size of a local area or, which is practically the same, the number of the neighbors.

The method of LA [4] historically has become the first local method developed for the forecast of time series on the basis of the Takens theory. This method also applies representation (2), but in this case it has three basic differences from a method of autoregression:

- coefficients in expression (2) are estimated separately for each of local areas;
- expression (2) can include also nonlinear members, i.e. various degrees of values of the series in the previous time (that is said to be an approximation order);

- embedding dimension is used as a value of the autoregression order.

Thus, the LA method turns out to be more valid for the choice "of the autoregression order" and more flexible in use of the initial data.

Let us stop briefly on the basic steps of algorithm of the LA method.

1. The choice of local representation

This step includes evaluation of the embedding dimension and construction of many-dimensional representation of a series, i.e. the matrix of delays X .

2. Determination of a vicinity, i.e. the number of neighbors

As a rule, the vicinity is given by a choice of number of the neighbors N^s . For this purpose in the state space (among the columns of the matrix X) the most "close" states are chosen. As the most simple criterion of the closeness it is possible to choose the following: for the given metric $\| \cdot \|$ and for *the given* number of neighbors N^s the set $\{y^t\}$ will be a set of the nearest neighbors x , if $\sum_{t=1}^{N^s} \|y^t - x\| \rightarrow \min$. It should be noted that the closeness of y^t to x , though we consider the change dynamics in the observable, does not mean the closeness in time.

3. Choice of an approximation model and its identification

In this step, for the determined embedding dimension the order of approximation is chosen (usually either linear approximation or square-law approximation in the previous instants are used):

$$x_1^{t+T} = f^{t+T}(x^t) = P_q(x_1^t, x_2^t, \dots, x_m^t), \quad (3)$$

where $T = 1, 2, \dots, q$ are the polynomial degree, i.e. the approximation order. Therefore schemes of LA can be classified according to the approximating polynomial order. Coefficients of the model are estimated by a Singular Value Decomposition (SVD) [11].

In cases of strongly limited length of a series the approximation of a zero order is used. Then the "forecast" depends only on the hit in the concrete local area. Such a situation can be illustrated by the weather forecast [12].

4. Forecasting

The last step is the forecast for the next elements of the series. This is made on the assumption that the evolution of the last elements occurs according to the same law, as for other vectors from their local vicinity.

To make a forecast for a few steps forward, two basic ways of extrapolation [12] are used. In the first or *iterative* way for $T = 1$ the model parameters in (3) are estimated, and the further forecast, $T = 2, 3, \dots$, represents a sequence of iterations. This means that the predicted value is added to an initial series. Then, assuming that the obtained new vector

in the state space are in the same vicinity (that means that it evolves by the same law), the forecast for one step forward is made. And so on.

The second, an alternative way, is *direct* forecast. It consists of the estimating of parameters separately for each T . This method allows to make more exact forecast, since in this case there is no accumulation of an error on each step [13].

By development of a LA method it was supposed, that the accuracy of predictions is limited by the quality of approximation, which for an available data set is determined by a number of points. However in many cases the accuracy of a forecast can be limited as well by noise: even if we know precisely the equations, the noise narrows limits of the predictability. It brings an error in the determination of the initial conditions and smears trajectories. In [12] it is shown that the influence of noise on the quality of the forecast is very similar (by the consequences) to the errors in approximation. In LA the fatal error arising under influence of noise appears as a choice of neighbors. Thus, it is necessary to generalize LA and overcome this difficulty.

2.3 Singular spectral analysis

Firstly the SSA method was developed for extraction periodic and quasi-periodic components from time series. It was shown that this method can be used for the improvement of a signal-to-noise ratio [6]. Moreover, recently there appeared options for the extension of SSA resources allowing to make on this basis the forecast of further dynamics of a series. However, in the present paper this method is used only for suppression of noise.

The basic idea of a SSA method consists of the transformation of the matrix X by the algorithm close to a method of principal components (PC). The use of PC is the most important part of SSA which distinguish it from other methods of nonlinear dynamics applied for the analysis and the forecast of time series.

The main point of PC is a decrease of the dimension of initial space of the factors (in this paper this is the space of delays) by means of the passage to more "informative" variables (coordinates). As a result, new variables are called the principal components (PCs). This passage is carried out via orthogonal linear transformation.

In practice, in order to pass to PC it is necessary to calculate eigenvalues and eigenvectors of the matrix XX^T . The last ones are chosen as a new basis:

$$\Lambda = V (XX^T) V^T, \quad (4)$$

where $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & & \lambda_M \end{pmatrix}$ is a matrix of eigenvalues, $V_{M \times M}$ is a matrix of eigenvectors.

In addition, the matrix of PCs is $Z = VX$. At such a transformation, PCs represent certain sets of point projections of an initial set into eigenvectors. The eigenvalues characterize the scatter of points along new axes. Usually they are ordered in the descending order.

The PCs have many useful properties. In SSA the resulting decomposition is used for the extraction of the most significant components and the truncation of random perturbations in the investigated series. The basic idea of such a filtration is the use for reconstruction not all PCs of a matrix X but only the most significant ones. The significance of PCs is usually determined by the values of their eigenvalues.

In general, approximation of a matrix X imply the following:

$$\hat{X} = V_{r \times r}^T V_{r \times r} X, \quad (5)$$

where $V_{r \times r}$ is a part of eigenvector matrix corresponding to r first principal components. After approximation of the matrix X it is necessary to reconstruct an initial time series. Reconstruction of the matrix X by the first principal components leads to the lost of the initial diagonal image. That is the reason why at the reconstruction of the initial time series it is necessary to average the matrix elements along diagonals with the originally identical values (sf. (1)):

$$\hat{f}_t = \begin{cases} \frac{1}{t} \sum_{i=1}^t \hat{x}_{M-t+i,t}, & t < M, \\ \frac{1}{M} \sum_{i=1}^M \hat{x}_{M-t+i,t}, & M \leq t \leq N - M + 1, \\ \frac{1}{N-t+1} \sum_{i=1}^{N-t+1} \hat{x}_{i,t-M+i}, & N - M + 1 < t \leq N, \end{cases} \quad (6)$$

Thus, the algorithm of reconstruction of time series by SSA includes three basic steps:

1. Construction of the matrix X .
2. Calculation of principal components and choice from them the most significant.
3. Reconstruction of the time series by the chosen principal components.

The application of this algorithm allows to smooth an initial series, decrease the noise level and raise a signal-to-noise ratio. However, the methods of forecasting [7] developed on its basis are insufficiently effective for unperiodical time series. Like the LA method, SSA has some modifications related to a preliminary centering and/or normalization of rows in

the matrix X . We shall use the variant with the centered rows because it is more close to the PC standard.

3 Generalization of LA by SSA. SSA-LA method.

In this section we present a certain generalization of the described methods via unification of some possibilities of LA and SSA.

3.1 Algorithm of SSA-LA

In the case of highly noisy series increasing the number of observation does not allow to apply LA algorithm effectively because it is highly probably that false neighbors will appear and, simultaneously true neighbors will be eliminated. Therefore, to improve the quality of the forecast it is necessary to combine both analyzed methods (SSA and LA). In this case SSA will be used only for the filtration of an initial time series (i.e. noise reduction), and the forecast will be made according to the LA method. Thus, the proposed method (SSA-LA) can be considered as a generalization of LA which allows to process highly noisy time series.

The SSA-LA algorithm consists of two basic stages:

1. SSA-filtration for noise reduction;
2. Forecasting of the modified time series by the LA method.

It should be noted that to make a forecast a LA variant of the first order and SSA with a centering were chosen. To estimate the coefficients in (2) the PC method was used instead of SVD (see [13]). This allows to raise the accuracy of the forecasts.

Application of SSA and LA shows that at the SSA reconstruction the first and the last M elements of the series can be found with quite large deviations from initial values, whereas other parts of the series are approximated much more precisely. Apparently, decreasing the accuracy of boundary value approximations is a result of a shorter interval of the averaging used in (6). Therefore for more qualitative forecasting the first and the last M values were truncated, i.e. the forecast was made starting from $N - M + 1$ value.

3.2 Numerical results

In this section some numerical results are presented. They can help to estimate the opportunities of SSA-LA application for the forecast of highly noisy irregular time series. These time series were generated on the basis of the known Mackey-Glass equation with

a delay (7) and x -components of finite-difference approximation of the Lorenz system (8). Uncorrelated Gauss noise with zero expectation were added.

$$x_{i+1} = (1 - b) \cdot x_i + \frac{a \cdot x_{i-\Delta T}}{1 + (x_{i-\Delta T})^c}, \quad a = 0.2, \quad b = 0.1, \quad c = 10, \quad \Delta T = 17, \quad (7)$$

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = (r - z)x - y \\ \dot{z} = xy - bz \end{cases}, \quad \sigma = 5, \quad r = 15, \quad b = 1, \quad \Delta t = 0.02. \quad (8)$$

It should be noted that base time series were designed once for all numerical analysis, and a noise component was generated independently for each experiment.

The results of numerical analysis are given in Figs. 1,2,3. For the forecast, 3600 points of each time series were chosen, and 110 points from them were used only for an estimation of the forecast quality. Parameters of a SSA-filtration are the following: $M = 18, r = 6$. LA parameters are: $m = 6, N^s = 21$.

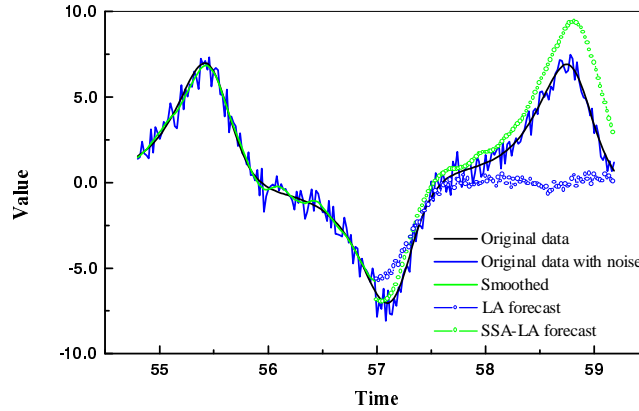


Figure 1: SSA-filtering and forecast for x -component for the Lorenz system with 5% noise. The forecast start point is 57.0.

First let us consider the example of x -component forecast for the Lorenz system (Fig. 1). The left part of the diagram (up to the point 57.0) shows the result of a SSA-filtration of a noisy signal (last 110 points), which is a quite accurate reconstruction of the initial series (without noise). However, the phase is reconstructed with a small enough error. From the point 57.0 the forecast begins (circles). One can see that the preliminary filtration allows to raise essentially the accuracy of the forecast. However basically the accuracy of the forecast can depend on the moment of the beginning. To get more real estimation of the forecast quality by LA and SSA-LA it is necessary to apply the characteristic which is not depends

on the start point. Following [12] such a characteristic is the value of the standard root-mean square error or the prediction error:

$$E = \sqrt{\frac{\left\langle (f_{t+T} - \hat{f}_{t+T})^2 \right\rangle_t}{\left\langle (f_t - \langle f_t \rangle_t)^2 \right\rangle_t}}. \quad (9)$$

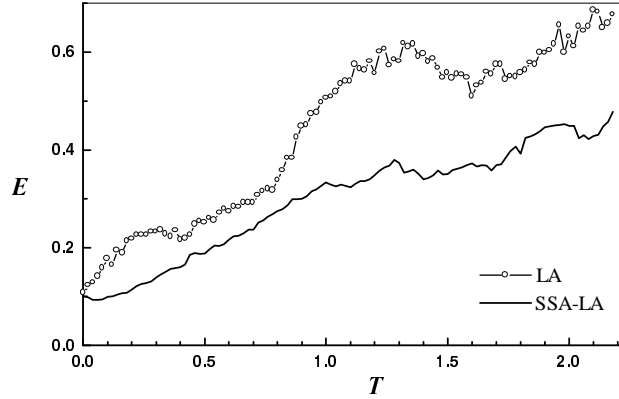


Figure 2: Prediction error for x -component of the Lorenz system with 5% noise. Averaging over 500 forecasts.

Here the averaging is made by the moments of the beginning of the forecast. If E exceeds one then the forecast is not successful, and would be better to exploit the average value as the forecast. For more reliability of the error estimation, expression (9) contains median instead of the arithmetic mean for the initial moments of the forecast.

For the same time series, Fig. 2 shows the prediction error as a function of the forecast length. Like the example in Fig. 1, the mean forecast with a preliminary filtration is better. The error in the SSA-LA case is, as a rule, less than the error for the standard LA. Dependencies shown in Fig. 2 were constructed for various amplitudes of noise. Except for zero noise level, the forecast was more exact for the LA-SSA method.

By the same criterion, the series obtained from the Mackey-Glass equation have been analyzed. The results turn out to be very similar to the Lorenz series Fig. 3. The observed advantage of the LA method in the absence of noise is a natural result of quite small distortions in the initial series obtained during the SSA-filtration. However for the noisy series this fact is completely compensated by suppression of noise. So, the forecast by the SSA-LA method is more precise than the standard LA-approach.

Thus, the numerical analysis show that the preliminary SSA-filtration allows us to raise considerably the accuracy of the forecast obtained by the LA method. In turn, combination

of LA-SSA can be very useful for the study and forecast the noisy time series when the standard LA is not efficient.

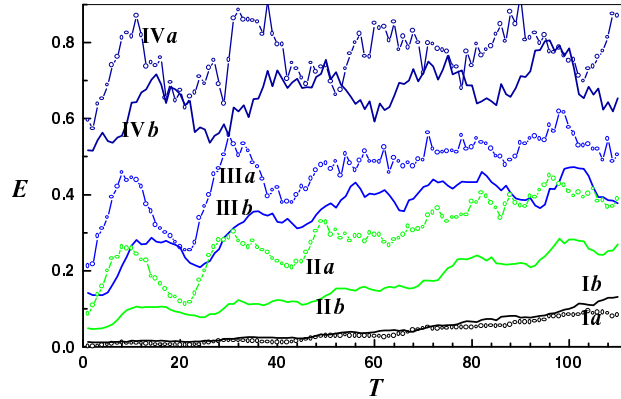


Figure 3: The prediction error for the Mackey-Glass equation as a function of the forecast length without noise (I) and with noise of the amplitude of 1.5% (II), 5% (III), 20%(IV). LA (a) and SSA-LA (b). Average over 500 forecasts.

4 Conclusion

The forecast of time series by the LA method is currently more an opportunity than a real researching tool. At the same time, the LA algorithm has a certain advantage over the usual autoregression. This advantage consists of the use of piecewise-linear approximation instead of a global-linear one. This allows to predict irregular time series for which a linear autoregressive representation can not be applied. The basic reason which limits the use of the LA method is that its efficient application is possible only for the forecast of a sufficiently long time series. In the case of a high noise level the requirements to the length of a series essentially grow; in quite rare situations one can find a series of the necessary length.

To decrease this limitation in the present paper we propose to use a preliminary filtration of a series by the SSA method. As known, SSA is a good tool for noise suppression especially for irregular time series, when the Fourier-filtration cannot be applied. It is numerically shown combination of SSA and LA gives more accurate forecast than in the standard LA method. This result practically did not depend on the noise level (except for the case of its absolute absence), on the length of the forecast and on the nature of the system which generated the analyzed series. It seems to be true that some advantage of LA over SSA-LA in the absence of noise is completely compensated by the significant advantage of the SSA-LA

for the noisy data because as a rule, real data includes a certain noise. Thus, one can say that the use of the SSA-filtration as a necessary component can essentially extend the area of the LA application.

References

- [1] Box G.E.P., Jenkins G.M., Time series analysis Forecasting and control, San Francisco, Holden Day, 1970.
- [2] Monin A.S., Piterbarg L.I., Weather and climate forecasting ability. In Kravtsov J.A. Forecasting limits, CentrCom, Moscow, 1997.
- [3] Takens F. Detecting strange attractors in turbulence, Lect. Notes in Math, Berlin: Springer, 1981, v.898, 336-381.
- [4] Farmer J.D. and Sidorowich J.J., *Predicting Chaotic Time Series*, Phys. Rev. Lett., 59 (1987), 845-848
- [5] Vautard R., Yiou P., Ghil M., Singular spectrum analysis: A toolkit for short, noisy chaotic signals. - *Physica D*, 1992, v.58, p.95-126.
- [6] Ghil M., Allen R.M., Dettinger M.D., Ide K., Kondrashov D., Mann M.E., Robertson A., Saunders A., Tian Y., Varadi F., and Yiou P., (2000), "Advanced spectral methods for climatic time series", Rev. Geophys., submitted.
- [7] Time series principle components: "Caterpillar" method. Digest by Danilov D.L. and Gillavskiy A.A.. - SPb university, 1997.
- [8] Broomhead D.S., King G.P., Extracting qualitative dynamics from experimental data. - *Physica D*, 1986, v.20, p.217-236.
- [9] Loskutov A., Istomin I.A., Kotlyarov O.L., *Testing and Forecasting the Time Series of the Solar Activity by Singular Spectrum Analysis*, Nonlinear Phenomena in Complex Systems, 4 (2001), 1, 47-57.
- [10] Grassberger P., Procaccia I., *Characterization of strange attractors*, Phys. Rev. Lett., **50**, (1983), 346-349
- [11] Kahaner D., Moler C., Nash S., Numerical Methods and Software, Prentice-Hill, 1989.
- [12] Farmer J. D. and Sidorowich J. J., *Predicting Chaotic Time Series*, Phys. Rev. Lett., **59** (1987), 845-848.
- [13] Loskutov A., Istomin I., Kotlyarov O., Some kinds of local approximation technique. in preparation.